DCCUMENT RESUME

ED 049 894 RE 003 497

AUTHCF Chester, Fobert
TITLE A Critical Examination of Vocabulary Control in View of Present Research Dealing with Initial Learning Tasks.
PUB DATE 5 Feb 71
NOTE 15p.; Paper presented at the meeting of the American Educational Research Association, New York, N.Y., Feb. 4-7, 1971

EDRS PRICE EDRS Price MF-$0.05 HC-$3.29
DESCRIPTORS *Associative Learning, *Beginning Reading, Cloze Procedure, Function Words, Kindergarten Children, *Readability, Reading Materials, *Reading Research, Semantics, Syntax, *Vocabulary, Word Frequency

ABSTRACT

Because vocabulary control in children's reading materials is largely based on vocabulary lists and studies which are now obsolete, new areas of learning are being considered for developing beginning reading materials. Instead of basing vocabulary lists on frequency counts and adult reading materials, the three approaches examined here are concerned with paired-associate learning which begins with the child. According to the concept of learnability, there are individual characteristics within printed words which make them easier or harder to learn than other words. A second approach to learning demonstrates that children's associations are predominantly syntagmatic in contrast to the paradigmatic associations of adults. Another area of learning related to children's vocabulary is the division of words into two categories: those which convey the message information and those which convey information about the structure, i.e., function words. Results of a recent pilot study with 24 6-year-old prereaders indicated that there is no difference in learnability between the two word classes, although previous research had reported a dichotomy existed. A bibliography is included. (DH)

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Paper - A.E.R.A.

February 5, 1971

Robert Chester

A Critical Examination of Vocabulary Control in View of Present Research Dealing with Initial Learning Tasks

ED049894

An examination of the vocabulary studies and lists which have in the past contributed large measures to vocabulary control reveals that most of them are poorly founded. Many of the lists which are influential today are obsolete or are based on lists and studies which are obsolete.

Probably the most important lists developed during the past fifty years are those of Thorndike (1921; 1932; 1944). Thorndike's lists are important, not only because of their influence on teaching materials and methods, but also because they have been used as one of the principle sources for most of the other vocabulary lists developed for more than a quarter century. A close examination of the Thorndike lists reveals several interesting things. First of all, the lists are frequency lists, i.e., they are lists based upon the frequency of the appearance of a word in context. The context used by Thorndike was that of material generally written for adults. For example, only thirteen percent of the material from which he took his first word list was written for children. There seems to be little in the sources which would recommend it as a developmental list for children's material. Some striking evidence in favor of this point may be found in one of Thorndike's own studies. He made a study (1936) of four and a half million running words from books recommended for elementary children. Of the 20,000 words in his Teacher's Word List (1932), only 2,500 occurred frequently. He also noted the occurrence of 18,000 words which were not on his 20,000 word list.

RE003 497

A second point, which though obvious is nevertheless important, is the fact that all of these frequency lists are based on written sources. The assumption that written material approximates the oral language vocabularies of children is no more than an assumption. There is no evidence to indicate that the lists were based on anything beyond frequency in written context.

Gates' word list (1935) avoided the criticism of being based on adult vocabulary context. However, it also has its limitations. It too, like Thorndike's lists, is based solely on frequency. Except for Horn's limited study (1924) of children's spoken vocabulary, it contains only sources based on written context. In this case, Gates used material from studies of children's literature, primary reading texts, and Thorndike's 1921 study. There is no indication that Gates attempted to verify his word list by checking it against the spoken vocabulary of children in the schools at the time of its publication.

Ernest Horn's study (1928) was second only to Thorndike's in extensiveness. Although this word list of more than five million running words was quite influential in the late 1920's and 1930's, it has less basis for validity than either the Gates or the Thorndike lists. It was based on business and personal letters of adults, and there is little justification for relating it to children's material. Unfortunately, it was used as a primary source for the Horn-Ashbaugh Fundamentals of Spelling (1928), and for the development of other word lists.

The three most influential frequency counts concerned with children's spoken vocabulary appear to be those of Horn (1924), Packer (1921), and Madeline Horn (1928). However, the first two studies mentioned failed to

describe the way in which the words were solicited, and all three studies fail to provide a description of their population which would meet even minimal standards. Nothing is given concerning the race, socio-economic status, background, I.Q., or other relevant circumstances. Consequently, it is impossible to generalize from these studies. Unfortunately, such generalizations were made since all three lists became primary sources in various other studies.

At the time of its inception, the Dolch list (1936) had little justification for acceptance. Although it has since exerted a very important influence on vocabulary development, its own basis is very unstable. The Madeline Horn list, one of its sources, was unsuitable for generalization to another population. The Gates' list (1936), another source, was based on the questionable criteria of the Thorndike list (1921), the Horn study (1924), and various other frequency studies based on written context. The third source of the Dolch list, the Wheeler and Howell study (1930), was merely a frequency listing of what was present in a number of basal vocabulary lists. Dolch did not base his list on children's vocabulary. Rather, he based it on materials which existed and on lists which were unsuitable for generalization.

Similarly, a look at the sources of the Buckingham and Dolch list (1936) reveals the same questionable foundation. Most of the sources are poorly substantiated and are really unsuitable for generalization to the oral language of children. Buckingham and Dolch attempted to remedy this by adding their list of free association words. However, since this list only constitutes a small portion of the source, and since it seems questionable that words solicited by stimuli are really free response words,

their remedy seems inadequate.

The Faucett and Maki frequency list (1932), based on the frequency lists of Thorndike (1921) and Horn (1928), appears to have the same limitations as the parent lists, only compounded. The same appears to be true of the Durrell frequency list (1936) which had as its sources the frequency lists of Faucett and Maki (1932) and Fitzgerald (1934).

One formidable vocabulary list, not based on previous studies, was that of Rinsland (1945). However, the Rinsland list has serious limitations of its own. Rinsland wrote to schools all over the United States to get samples of children's compositions. He received only 47% return on his request. This seems adequate in view of the fact that this return consisted of over 200,000 samples from 708 schools. However, when we examine the sources, an obvious bias is present. New England states were represented by 50 schools, the middle Atlantic states by 91 schools, the northwestern states by 43 schools, the southwestern states by 62 schools, the southern states by 250 schools and the central states were represented by 212 schools. These figures indicate that approximately 66% of his entire population came from the southern and central states. One also notes that 185 of the 236 county schools came from these same areas.

Besides having a biased population, Rinsland had other problems. He notes that after eliminating samples which were duplications, or which were questionable as to validity, he had only 103,000 left. He neglects the geographical locations from which he threw out half his sample population. The frequency and grade level list which he compiled from the remainder of the return seems questionable, to say the least.

A brief survey of some of the major word lists and vocabulary studies of the past, although admittedly inadequate, does yield some basis for several conclusions. first, the attribute which all of the lists appear to have in common is that they are all frequency counts, and, for the most part, they are based on other frequency counts. In some cases, Durrell for example, they are lists based upon lists which are based upon still other lists. Secondly, they are generally based upon written or text-book samples and not on the oral language of children. Much of the material was written at adult levels or came from adult concepts of children's language. Textbooks written for children, based only upon the adult's intuition together with questionable word lists seems to leave something to be desired. Possibly the most damaging limitation of the studies and lists lies in the fact that they do not go to the primary source, the children. None of the word lists were verified in the classroom before their publication. Generally, the actual language of children was not considered as a primary source. In the few cases where the oral vocabulary lists were considered and used, their populations were so porly reported that generalization from them was unwise.

Up until the present, vocabulary control has been based primarily on the frequency of occurrence of the words in the language. Although a number of educators have indicated that this is an inadequate criteria, very little has been done about the situation. This paper will examine three relatively new areas of learning for consideration in developing

vocabularies for beginning readers. It is not incidental that all three of these approaches are based on paired associate learning. Underwood and Schulz (1960) have shown that paired-associate learning can be analyzed into three stages: (a) discriminating the stimuli from one another, (b) making the responses available in the learner's repertoire, (c) pairing the stimuli and responses appropriately. Although it may be recognized that this is only one of several possible ways of examining learning, it should be noted that this one begins with the child.

Learnability:

The learnability construct is derived from Coleman (1968). He has shown that the frequency of occurrence of words in the language is poorly correlated with what he calls the "learnability" of words when learners are first grade children who are pre-readers. There are individual characteristics within printed words which make them easier or harder to learn than other words. Although these characteristics have not all been properly defined, they do seem to be present. Thus some words are more learnable than others in that their mastery is more easily accomplished by beginning readers. Coleman obtained a learnability measure by using as stimuli the 500 most common words in English according to the Lorge magazine count and the Lorge-Thorndike semantic count. The learnability scale was based on the mean number of misses in individual teaching-testing treatments: the more difficult the word, the greater number of misses in learning to recognize and respond to it.

Studies using similar treatments and yielding similar results have been carried out by Jones (1968) and Bickley (1969).

It is important to note here that the theory of learnability has been substantiated in studies involving concepts other than words. Using a paired-associate task, Bridge (1968) performed an experiment that rank-orders 35 letters and letter combinations according to the ease with which children learn their sounds. Laumback (1968) rank-ordered 293 two-sound words according to their phonic blendability and found that major differences, some as high as ten to one, existed in the different phoneme combinations. Coleman (in press) replicated her study and verified her findings. He suggests that the results of these studies are applicable for most children. A. Jones (1968) did a study in which she rank-ordered the lower-case letters according to ease of printing. From the results of this study she was able to generate tables and learning curves for the three common errors of the letters of the alphabet.

In light of these studies and others like them which are appearing more and more frequently in educational research, it seems imperative that vocabularies for beginning readers take into consideration the concept of learnability. It seems apparent that rank-orderings of various information and concepts can be devised which may aid beginning readers.

Paradigmatic - Syntagmatic relationships:

One relatively new concept appearing in reading research is that of the paradigmatic and syntagmatic relationships among words. The designations paradigmatic and syntagmatic seem to have been proposed by Ervin (1957). She indicated that paradigmatic associates are applied to responses which are elicited by stimulus words of their own grammatical part of speech

(or form class) in the free-associates situation, while syntagmatic associates are responses which are of different parts of speech from the stimulus words and which seem to be part of the natural language sequence. In response to the stimulus word Hot, one might get the paradigmatic response of Cold, whereas the syntagmatic response might be Stove.

In investigating the categorization methods of individuals, it immediately becomes apparent that children and adults respond differently. Children's associations differ from adult associations. At younger ages, children's associations are predominantly syntagmatic. Adults, on the other hand, tend to give predominantly paradigmatic association responses. Rather than responding to the word dark with the light, the child tends to respond with night. (Entwisle, 1966; Palermo and Jenkins, 1964)

Bickley (1969), in his study with first graders, found significant differences between the learning of lists organized by logical congruence (paradigmatic associates) and lists organized by the typical associates (syntagmatic associates) as given by children. He suggests that the internal organization of associations in children is more congruent with sequential categorization than with semantic categorization. The associations of children more often than not conform to the syntax of the natural language.

An important concept which now emerges is the fact that children seem to group words by syntax, while adults group the same words by what appears to be a combination of the same part of speech and a logical semantic characteristic of the concept underlying the word. This difference is not a trivial one when notes its implication that underlying the child's categorizations is a lack of semantic organizational depth. Not only does

this have far reaching implications for the construction of basic learning materials, it is vitally important in the construction of evaluation instruments. (Bickley and Dinnan, in Press).

Syntactical Structural Units:

Generally, one finds that all utterances contain two types of elements: those that convey the message information and those that convey information about the structure of the utterance. In Latin and Greek, the root of the word usually conveyed the message (serv-: "slave"), and the endings conveyed the structural information (-us: masculine, singular, subject of sentence) or (-o: masculine, singular, indirect object). In modern languages these elements are often separated into distinct words: some of the words in the sentence convey the message information, others have the main purpose of conveying information about the structure. These latter may be called "function" words. There are only about 300 of them in our over 600,000 word dictionary and most of these appear in the first 1000 words selected by frequency of usage. Approximately 46% of the Dolch list is made up of these function words as is 43% of the Fry list. Since the Dolch and Fry lists claim to make up about 65% of the vocabularies for readers for the first three grades, one can see that these words are of extreme importance in beginning reading.

Despite the crucial role of "function" words, both because of frequency of appearance and because of structural usage, most compilers or teachers do not differentiate between them and content words. Fry himself admits the difficulty of teaching them as vocabulary items because they are "largely

devoid of subject-matter meaning or object reference." (Fry, 1960). Taylor, (1957) conducting a study in which he deleted words which corresponded to lexical and structural words, found that there was a significant difference in scores for the two types of word structures. Whereas he found that lexical cloze was a significantly better predictor of achievement, he found structural cloze to be a better measure of readability. Jefferson (1968) also found a difference in the ability of both high and low ability readers to supply structural function and lexical (content) items deleted in a cloze test. Weaver (1964) even goes so far as to suggest that the two categories of words are stored separately within the cognitive structure. There seems to be ample justification for believing that a dichotomy does exist.

Druing the summer of 1970, the writer carried out a pilot study with six year old pre-readers who were entering school for the first time in September. The subjects were 12 high and 12 low socio-economic kindergarten or headstart children from Clarke County, Ga. Using a paired-associate task similar to that of Coleman, the examiner attempted to establish the ease of learning, or learnability, of content and function words being used as stimulus.

An analysis of the data collected in the study revealed several interesting things. First, the examiner found no significant difference in the learnability scores of subjects for content or function words. This was somewhat surprising in view of the literature on the differences between the two word classes.

Secondly, the examiner found that fifteen trials were not sufficient for mastery of the words in either class. Scores for both word classes were higher than had been anticipated. This weakness in procedure was eliminated

in later testing; however, it was discovered too late to be rectified in the pilot study. The combination of scores that were too high with a ceiling that was too low had serious effects on the analysis of the data.

A third discovery indicated that almost half of the low socio-economic subjects failed to "learn" the stimulus words. This was true of only one high socio-economic subject. Unfortunately the effect of the low ceiling prevented the scores of the low socio-economic students from rising. Consequently, it was again impossible to establish any learning differences.

Despite the obvious failure of the pilot study to provide any significant data, the examiner found it invaluable in planning the more extensive study carried out in the fall and winter of 1970.

By improving the procedures, providing better controls and raising the ceiling, the examiner was able to develop a more realistic picture of the situation concerning the two word classes. The data from the major study is now being analyzed and studied. Preliminary analysis of the data reveals that the differences between content and functions words are not present. Although the examiner plans further analysis of the data before confirming this, he tends to believe this may be the case. The lack of significant differences, if it does exist, is again somewhat surprising in view of the literature.

In view of this possibility of no difference, the writer offers two explanations, both of which are speculation. First, since all of the previous studies dealing with content and function words have been carried out with older subjects, mostly adult populations, the differences may be learned or developed at a later period. In view of this, the writer is

planning a replication of the original study with an older population.

The second speculation is that the differences in the methods used in examining content and function words may have something to do with the differences in results. That is to say, the cognitive processes used in the paired-associate task may be different from those in the clozed task. In either case, it is apparent that further investigation is necessary before any difinite conclusions can be reached.

In conclusion, we must go back to the original point that we can no longer assume that frequency is an adequate control in developing vocabulary lists for beginning readers. Current research shows us that beginning readers do not learn in the same way as adults; perhaps not even in the same way as second grade pupils. By taking advantage of what we already know about children's learning patterns and by continued research and application of research, we should be able to better educate children in today's schools.

References

Bickley, A. C. An analysis of frequency, learning rate, and semantic congruence as categorization constraints on beginning readers. Unpublished doctoral dissertation, University of Georgia, 1969.

Bickley, A. C., Dinnan, J., and Bickley R. Reading achievement compared with oral language response, Reading Teacher, (in Press).

Bridge, J. T. Rank-ordering of letters and letter combinations according to ease of learning their sound associations. Unpublished master's thesis, University of Texas at El Paso, 1968.

Buckingham, B. R., & Dolch, E. W. A combined word list. Boston: Ginn, 1936.

Coleman, E. B. Collecting a data base for a reading technology. Journal of Educational Psychology, in Press.

Dolch, E. W. A basic sight vocabulary. Elementary English Journal. 1936, 36-37, 566; 268-272.

Durrell, D. D. A vocabulary for corrective reading. Elementary English, 1934, 11, 106-109.

Entwisle, D. Form class and children's word associations. Journal of Verbal Learning and Verbal Behavior, 1966a, 5, 558-565.

Ervin, S. M. Grammar and classification. Paper read at American Psychological Association, New York, September, 1957.

Faucett, L., & Maki, I. A. A study of english word values. New York: Oxford University Press, 1932.

Fitzgerald, J. A. The vocabulary of children's letters written in life. Elementary School Journal, 1934, 34, 358-370.

Fry, E. B. Three hundred instant words. Elementary English, 1960, 37, 38-42.

Gates, A. I. A reading vocabulary for the primary grades. New York: Teachers College, Columbia University, 1935.

Horn, E. The commonest words in the spoken vocabulary of children up to and including six years of age. The twenty-fourth yearbook of the National Society for the Study of Education. Bloomington: Public School Publishing Co., 1925.

Horn, E., & Ashbaugh, E. J. The Horn-Ashbaugh fundamentals of spelling. Philadelphia: J. B. Lippincott, 1928.

Horn, M. D. A study of the vocabulary of children before entering the first grade. Baltimore: International Kindergarten Union, 1928.

Jefferson, G. L. Lexical and structural items as predictors of readability for high and low ability readers. Unpublished doctoral dissertation, University of Georgia, 1969.

Jones, A. S. Rank-ordering of letters according to ease of printing. Unpublished master's thesis, University of Texas at El Paso, 1968.

Jones, H. D. A rank-ordering of common words according to their ease of look-say learning. Unpublished master's thesis, University of Texas at El Paso, 1968.

Laumback, J. D. Rank-ordering two-sound words as to phonic blendability. Unpublished master's thesis, University of Texas at El Paso, 1968.

Packer, J. L. The vocabularies of ten first grade readers. The twentieth yearbook of the National Society for the Study of Education. Bloomington: Public School Publishing Co., 1921.

Palermo, D. S., and Jenkins, J. J. Word association norms grade school through college. Minneapolis: University of Minnesota Press, 1964.

Rinsland, H. D. A basic vocabulary of elementary school children. New York: MacMillan, 1945.

Taylor, W. L. Cloze readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology, 1957, 41, 19-26.

Thorndike, E. L. Teacher's word book. New York: Teachers College, Columbia University, 1921.

Thorndike, E. L. Teacher's word book of 20,000 words. New York: Teachers College, Columbia University, 1932.

Thorndike, E. L. The vocabulary of books for children in grades 3 to 8. Teachers College Record, 1936-37, 38, 196-205.

Thorndike, E. L., & Lorge I. The teachers word book of 30,000 words. New York: Teachers College, Columbia University, 1944.

Underwood, B. J., and Shulz, R. W. Meaningfulness and verbal learning. Chicago: J. B. Lippincott Co., 1960.

Weaver, W. W. On the Psychology of reading. In E. L. Thurston and L. E. Hafner (Eds.), Thirteenth Yearbook of the National Reading Conference, Milwaukee: National Reading Conference, 1964.

Wheeler, H. E., & Howell, E. A. A first grade vocabulary. Elementary School Journal, 1930, 31, 52-60.